December 18, 2017

Omni-Directional Catadioptric Acquisition System

Andreas Nowatzyk

Andrew I. Russell

Omni-Directional Catadioptric Acquisition System

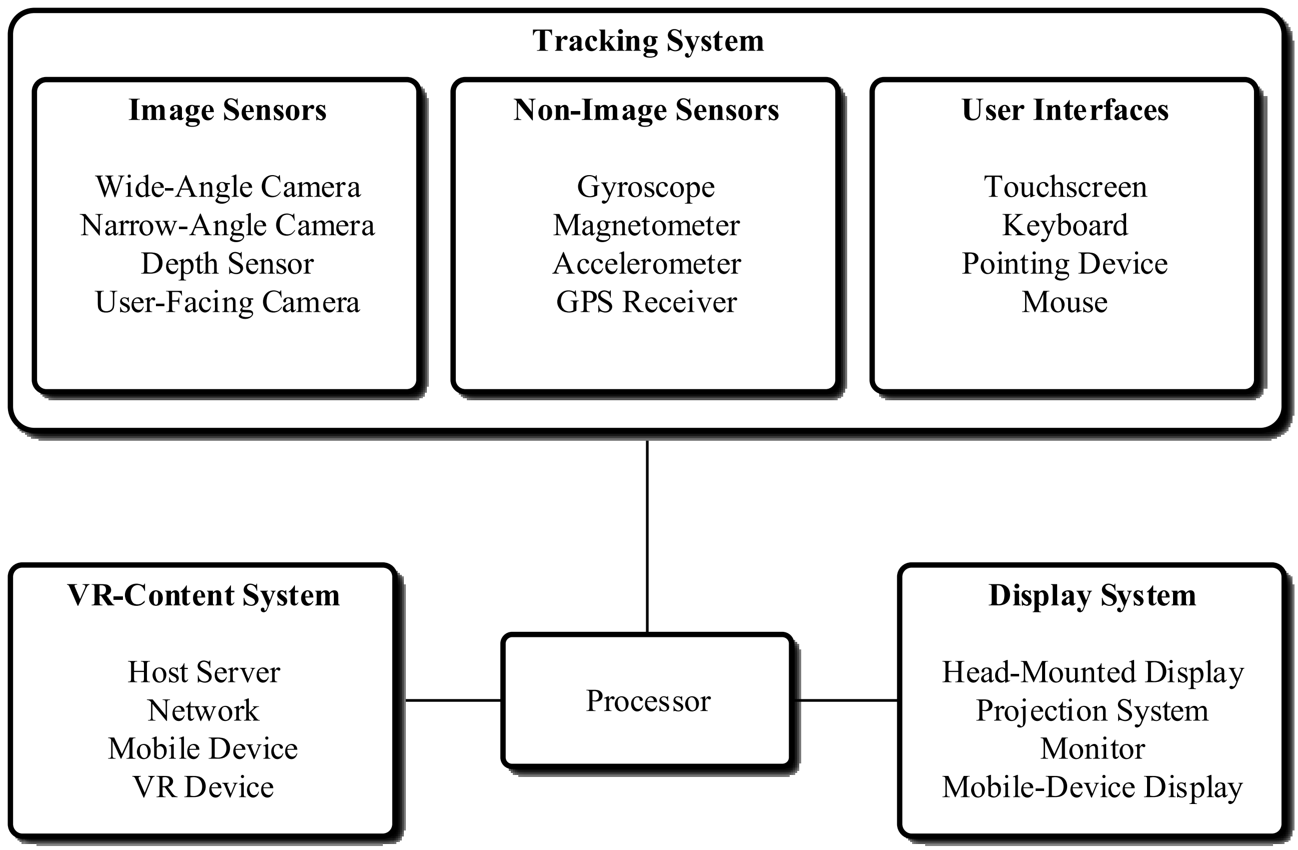
Abstract:

An omni-directional catadioptric acquisition system (ODCA system) is provided to address the problem of producing real time, 360°, stereoscopic video of remote events for virtual reality (VR) viewing. The ODCA system is a video image-capture assembly that includes a cylinder with multiple apertures arranged around its circumference to admit light as the ODCA system rotates about a central axis. Inside the cylinder, there is a mirror on the left and right side of each aperture that reflects light rays into the cylinder from different angles. As the cylinder rotates, the light rays are admitted through the apertures and reflected from the two mirrors to a curved mirror in the center of the cylinder. This curved mirror directs the rays down through a catadioptric lens assembly, which focuses the rays onto another curved mirror near the bottom of the ODCA system. This second mirror reflects the rays to a set of line-scan image sensors arranged around the second mirror. The line-scan image sensors capture the rays for later reproduction as stereoscopic video.

Keywords: virtual reality, VR, real-time video, VR video, remote location video, 360° video, 360 degree video, stereoscopic video, catadioptric lens

Background:

Virtual reality (VR) environments rely on display, tracking, and VR-content systems. Through these systems, realistic images, sounds, and sometimes other sensations simulate a user's physical presence in an artificial environment. Each of these three systems are illustrated below in Fig. 1.



VR System

Fig. 1

The systems described in Fig. 1 may be implemented in one or more of various computing devices that can support VR applications, such as servers, desktop computers, VR goggles, computing spectacles, laptops, or mobile devices. These devices include a processor that can manage, control, and coordinate operations of the display, tracking, and VR-content systems. The devices also include memory and interfaces. These interfaces connect the memory with the systems using various buses and other connection methods as appropriate.

The display system enables a user to "look around" within the virtual world. The display system can include a head-mounted display, a projection system within a virtual-reality room, a monitor, or a mobile device's display, either held by a user or placed in a head-mounted device.

The VR-content system provides content that defines the VR environment, such as images and sounds. The VR-content system provides the content using a host server, a network-based device, a mobile device, or a dedicated virtual reality device, to name a few.

The tracking system enables the user to interact with and navigate through the VR environment, using sensors and user interfaces. The sensors may include image sensors such as a wide-angle camera, a narrow-angle camera, a user-facing camera, and a depth sensor. Non-image sensors may also be used, including gyroscopes, magnetometers, accelerometers, GPS sensors, retina/pupil detectors, pressure sensors, biometric sensors, temperature sensors, humidity sensors, optical or radio-frequency sensors that track the user's location or movement (*e.g.*, user's fingers, arms, or body), and ambient light sensors. The sensors can be used to create and maintain virtual environments, integrate "real world" features into the virtual environment, properly orient virtual objects (including those that represent real objects, such as a mouse or pointing device) in the virtual environment, and account for the user's body position and motion.

The user interfaces may be integrated with or connected to the computing device and enable the user to interact with the VR environment. The user interfaces may include a touchscreen, a keyboard, a pointing device, a mouse or trackball device, a joystick or other game controller, a camera, a microphone, or an audio device with user controls. The user interfaces allow a user to interact with the virtual environment by performing an action, which causes a corresponding action in the VR environment (*e.g.,* raising an arm, walking, or speaking).

The tracking system may also include output devices that provide visual, audio, or tactile feedback to the user (*e.g.*, vibration motors or coils, piezoelectric devices, electrostatic devices, LEDs, strobes, and speakers). For example, output devices may provide feedback in the form of blinking and/or flashing lights or strobes, audible alarms or other sounds, songs or other audio files, increased or decreased resistance of a control on a user interface device, or vibration of a physical component, such as a head-mounted display, a pointing device, or another user interface device.

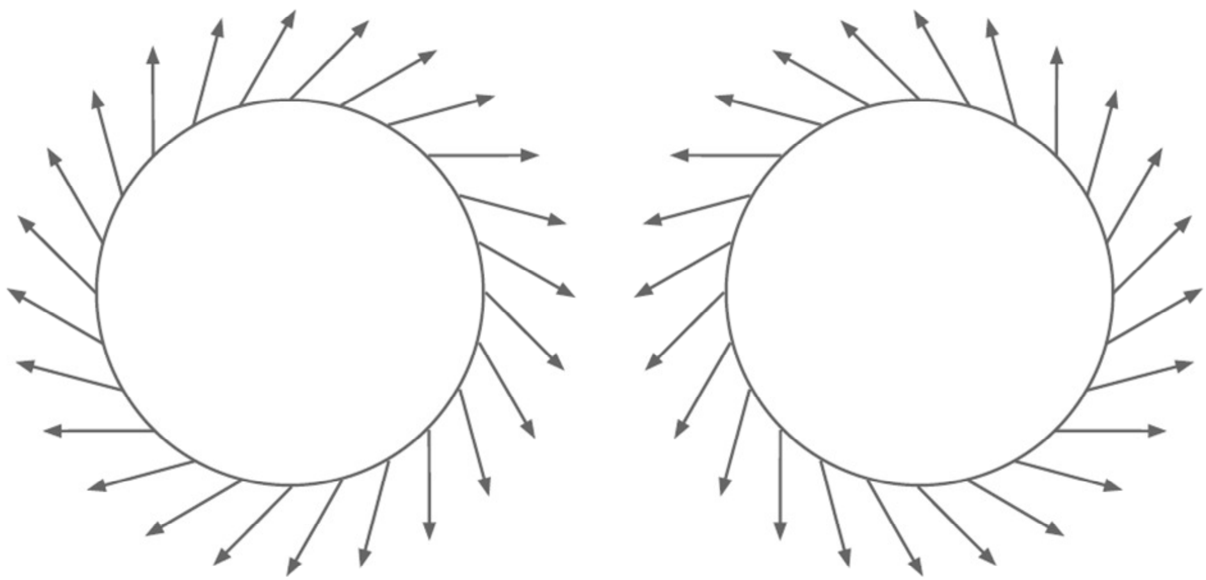
Fig. 1 illustrates the display, tracking, and VR-content systems as disparate entities in part to show the communications between them, though they may be integrated, *e.g.*, a smartphone mounted in a VR receiver, or operate separately in communication with other systems. These communications can be internal, wireless, or wired. Through these illustrated systems, a user can be immersed in a VR environment. While these illustrated systems are described in the VR context, they can be used, in whole or in part, to augment the physical world. This augmentation, called "augmented reality" or AR, includes audio, video, or images that overlay or are presented in combination with the real world or images of the real world. Examples include visual or audio overlays to computing spectacles (*e.g.,* some real world-VR world video games or information overlays to a real-time image on a mobile device) or an automobile's windshield (*e.g.,* a heads-up display) to name just a few possibilities.

Real time, 360°, stereoscopic video of remote events for virtual reality (VR) viewing is becoming a desirable part of a VR experience. This kind of video allows a viewer to view stereoscopic video with a 360° view from a fixed location. For example, the viewer can use a VR headset (*e.g.*, the head-mounted display described as part of the VR system of Fig. 1) to observe an event, such as a concert, an athletic competition, or a lecture, as though the viewer were present at the event. Providing this type of VR video reproduction can be difficult because the VR user's head position is unknown at the time a video is captured. Existing solutions use multiple video cameras that cover a sphere centered around the view point so that each point on the sphere is covered by at least two cameras with known baselines to allow three-dimensional (3D) image reconstruction. The resulting 3D model of the vicinity of the view-point is then used to synthesize the stereo-pair for the user's head attitude. This process, however, is computationally intensive, sometimes requiring more than 24 hours to process, which prevents this technique from providing real-time VR video.

Description:

To address the problem of producing real time, 360°, stereoscopic video of remote events for virtual reality (VR) viewing, an omni-directional catadioptric acquisition system (ODCA system) is provided. The ODCA system is a video image-capture assembly that includes a cylinder with multiple apertures arranged around its circumference to admit light as the ODCA system rotates about a central axis. Inside the cylinder, there is a mirror on the left and right side of each aperture that reflects light rays into the cylinder from different angles. As the cylinder rotates, the light rays are admitted through the apertures and reflected from the two mirrors to a curved receiving mirror in the center of the cylinder. The receiving mirror directs the rays down through a catadioptric lens assembly, which focuses the rays on a curved sensor mirror near the bottom of the ODCA system. The sensor mirror reflects the rays to a set of vertical line-scan image sensors arranged around the sensor mirror. The line-scan image sensors capture the rays for later reproduction as stereoscopic video.

As noted, the light rays that come through each aperture are admitted at different angles via the two mirrors inside the apertures, which allows the ODCA system's imaging software to create a stereoscopic image of the captured scene using the images from two different angles that are received by the line-scan sensors as a left view and a right view. Because the left and right views are directly captured by the ODCA system, the computational requirements are less than those of typical remote VR video systems. Stereoscopic video can be produced by stitching the left and right views together and using conventional image-processing techniques to produce a video of a remote event in near real-time. The video can be viewed with any of the displays described in the display system of Fig. 1 (*e.g.*, a head-mounted display or VR headset), allowing the viewer to have a stereoscopic, 360° view from a fixed position. Fig. 2 illustrates the 360° left and right views that are produced using the ODCA system.



Right-eye view from light rays deflected from one direction

Left-eye view from light rays deflected from another direction

Fig. 2

Fig. 3 illustrates an example configuration of the ODCA system. The example configuration includes an aperture cylinder, with nine narrow vertical apertures spaced around its circumference. A flat vertical aperture mirror is attached to each side of the aperture, inside the aperture cylinder. The aperture cylinder surrounds a curved receiving mirror that is custom-shaped for the particular physical configuration of the aperture cylinder (*e.g.*, diameter, height, number of apertures, and/or rotation speed). The aperture cylinder is described in additional detail with reference to Fig. 4.

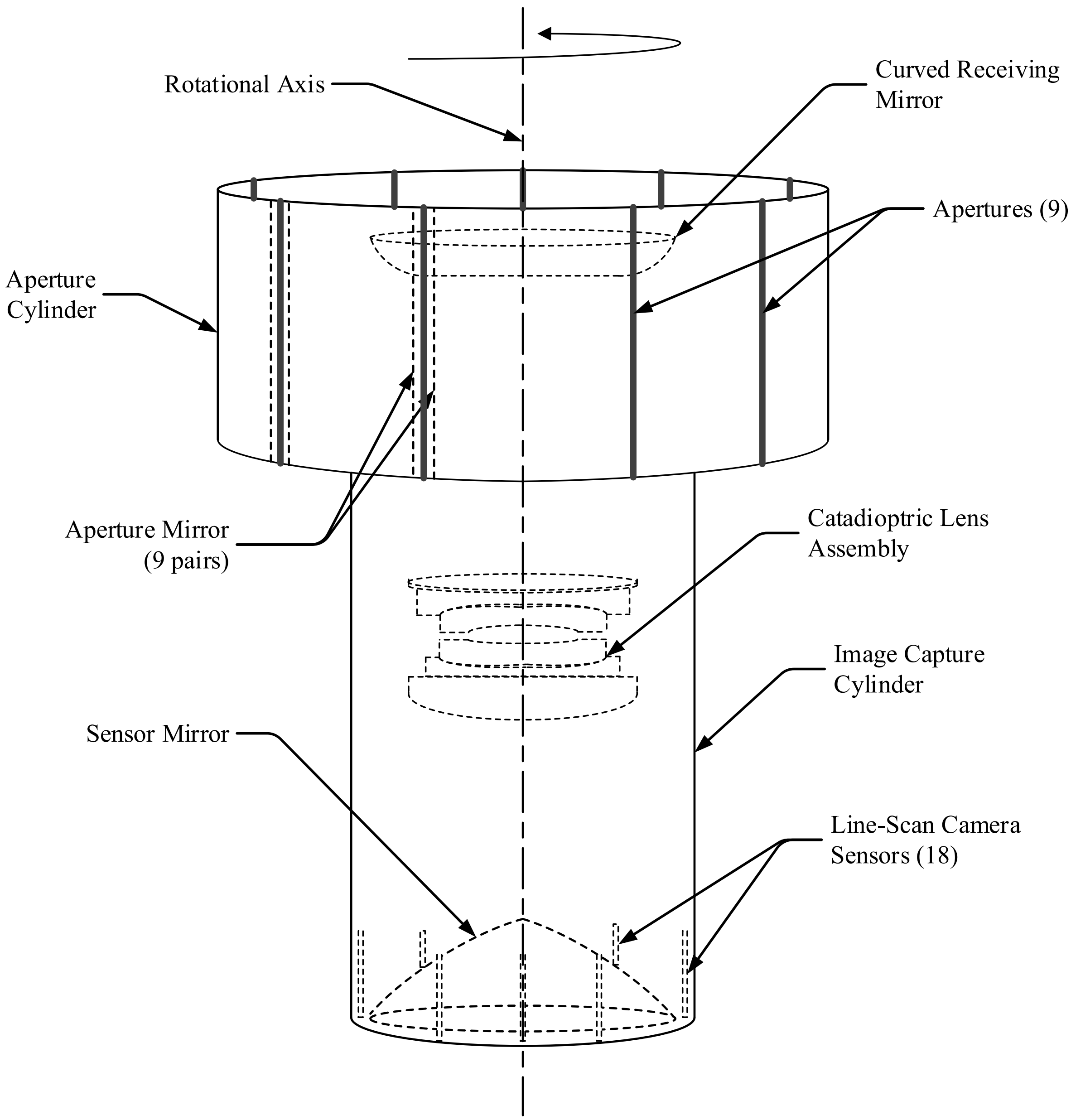


Fig. 3

The example configuration also includes an image capture cylinder that surrounds the receiving mirror and further modifies the path of the light rays received through the apertures (the light path is described in further detail with reference to Fig. 5). Inside the image capture cylinder, the light rays that are admitted through the apertures and reflected down from the receiving mirror are focused through a catadioptric lens assembly. The catadioptric lens uses reflection and/or refraction to direct the light rays to a sensor mirror near the bottom of the image capture cylinder. The sensor mirror is another curved mirror that is custom shaped to reflect the light rays received from the catadioptric lens assembly to a set of vertical line-scan image sensors that are arranged around the inside circumference of the image capture cylinder. In the example configuration of Fig. 3, there are 18 line-scan image sensors, one for each of the nine left-eye and nine right-eye views captured via the aperture mirrors.

As shown in the example configuration of Fig. 3, the receiving mirror and the sensor mirror are aspherical mirrors that are custom shaped to provide the proper reflection angles and focal lengths to direct the light rays representing the left-eye and right-eye views to the line-scan image sensors. The catadioptric lens assembly is a conventional catadioptric lens that shortens the length of the image capture cylinder that is required to properly orient the light rays to be directed off of the sensor mirror to the line-scan image sensors. The properties of the catadioptric lens assembly may vary with the design requirements of the ODCA system. The line-scan image sensors are conventional red, green, blue, and white (RGB+W) color line-scan sensors. Optionally, time delayed integration (TDI) line-scan sensors may be used, but because TDI sensors are generally monochrome, three times as many sensors are needed (one each for red, green, and blue), meaning for this example 54 sensors, as well as a separate RGB filter.

Fig. 4 illustrates a top view of the aperture cylinder depicted in Fig. 3. As shown in Fig. 4, the nine apertures are arranged symmetrically around the circumference of the aperture cylinder. The 18 aperture mirrors are attached inside the apertures, one on each side of the opening. The left light rays (red) and right light rays (green) are shown entering the aperture and being reflected off the aperture mirrors to the receiving mirror (solid-line circle, not labeled). Comparing the left and right light rays shown in Fig. 4 with the left and right views of Fig. 2 illustrates how the aperture cylinder captures the rays used to generate the 360° stereoscopic video.

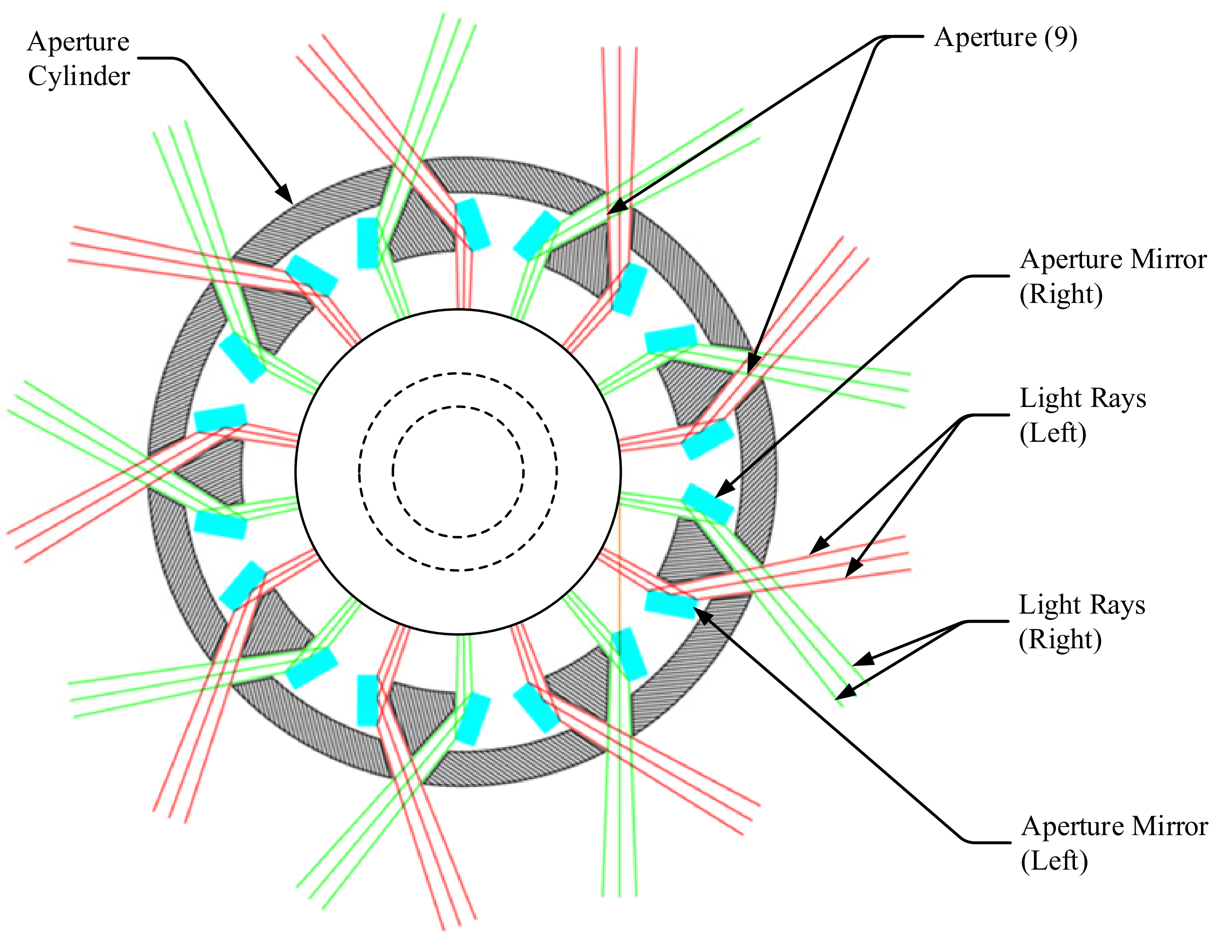


Fig. 4

Fig. 5 illustrates additional details of the light path as described with reference to Fig. 3. For clarity, the light path shown in Fig. 5 has been simplified to describe the concept. Thus, the figures show only two each of the left and the right light rays, two of the 18 aperture mirrors (one left, one right), and two of the 18 line-scan image sensors (one left, one right). Light from the scene being recorded is reflected off the aperture mirrors as the ODCA system rotates. Light rays representing the left-eye view (red) and light rays representing the right-eye view (green) are reflected from the aperture mirror to the receiving mirror, which directs the rays to the catadioptric lens assembly. The catadioptric lens assembly further directs the rays to the sensor mirror, which reflects the left and right light rays to the left and right line-scan image sensors, respectively. As described above, the line-scan image sensors receive the light and begin the image-processing sequence that produces the 360° stereoscopic video.

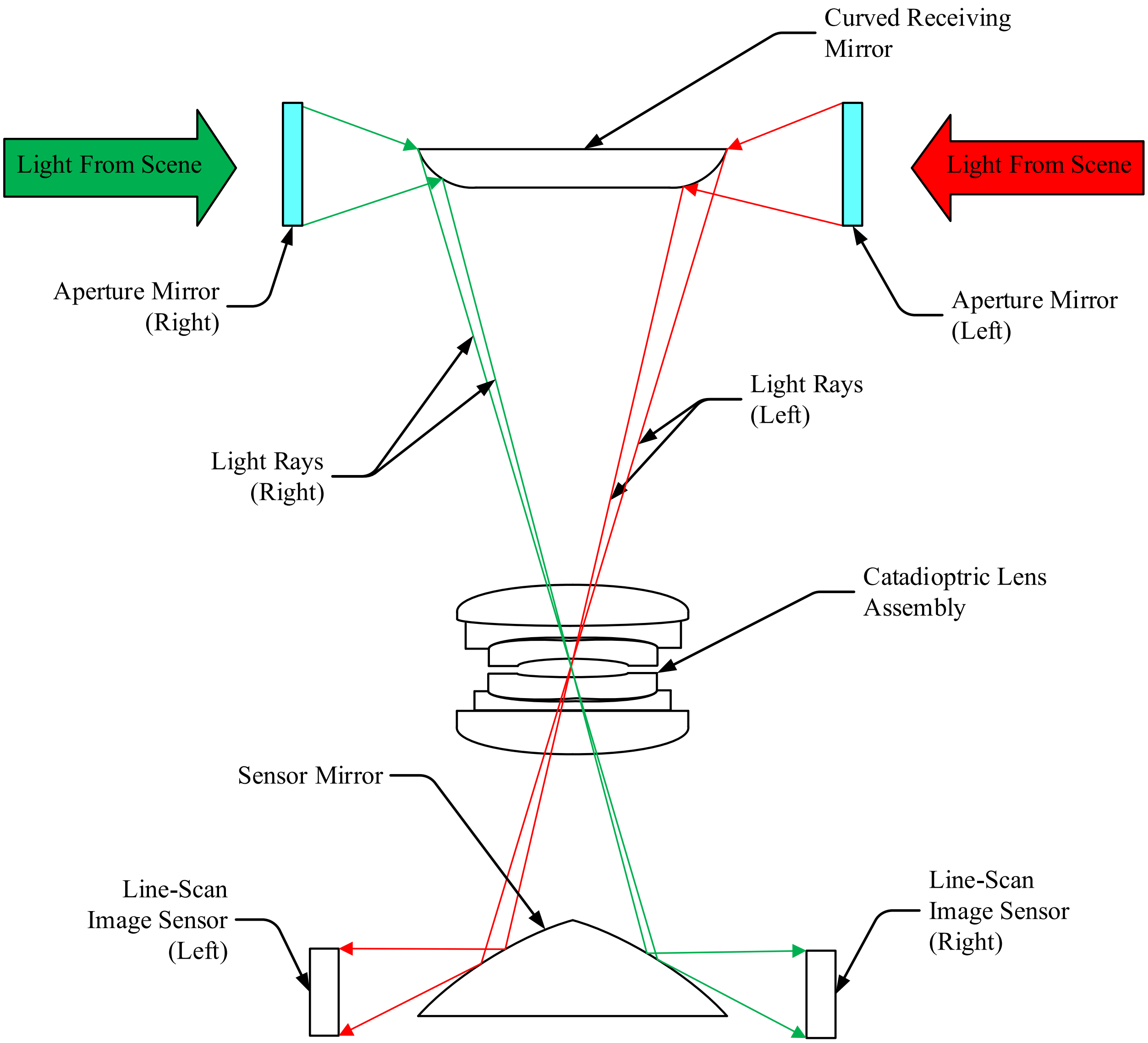


Fig. 5

The example configuration of the ODCA system described above, and shown in Figs. 3-5, includes nine apertures, 18 aperture mirrors, and 18 linear line-scan image sensors, nine to capture the images for the left eye and nine to capture the images for the right eye. The number of apertures and sensors is selected based on the relationship between the desired video resolution, the rotational speed of the ODCA system, and the size of the apertures. Smaller apertures may require a larger number of apertures. Further, each added aperture adds a corresponding sensor pair, which incurs costs related to the electronic components and increases the computational burden for video-processing. Conversely, while fewer sensors require faster rotation for a given resolution, exposure time per pixel (at the line-scan image sensors) decreases as the rotational speed increases.

Similarly, the optical design of the aspherical mirrors (the receiving mirror and the sensor mirror) is customized for the physical configuration of the ODCA system and the light path design as described above. The size and shape of the mirrors is related to the diameter of the aperture cylinder, which is generally determined based on the average eye separation distance for the viewers. Thus, other physical configurations of the ODCA system are also possible (*e.g.*, a different number of apertures and sensors, a different diameter and/or shape of the aspherical mirrors and the catadioptric lens assembly), depending on cost and performance tradeoffs related to the particular configuration and its design targets.